



## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/125148>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

# De voortzetting van de historische taalkunde met andere middelen

INAUGURELE REDE PROF. DR. NICOLINE VAN DER SIJS

## INAUGURELE REDE

PROF. DR. NICOLINE VAN DER SIJS



Digitale onderzoeksmethoden zijn binnen de geesteswetenschappen sterk in opkomst. De verwachtingen zijn hooggespannen, zowel bij onderzoekers als bij subsidiegevers. Terwijl de computer met moderne teksten al redelijk uit de voeten kan, zijn er voor oude teksten nog veel moeilijkheden te overwinnen.

In haar oratie laat Nicoline van der Sijs zien hoe digitale technologie ondanks beperkingen met succes kan worden gebruikt om oude en nieuwe historisch-taalkundige onderzoeksvragen te beantwoorden. Zo geeft ze antwoord op de oude vraag in hoeverre de gezaghebbende Bijbelvertaling uit 1637, de zogenoemde Statenvertaling, in taalgebruik volgend of leidend is geweest. De mythe wil dat de Statenvertaling grote invloed heeft uitgeoefend op de Nederlandse standaardtaal. Uit digitaal onderzoek blijkt echter dat de Statenvertaling bij verschijnen al archaïsche trekken had. Haar conclusie is dat historische taalkunde en digitale onderzoeksmethoden een gelukkige combinatie vormen: de nieuwe onderzoeksinstrumenten werpen vaak een verrassend nieuw licht op oude, fundamentele vragen rond taalverandering.

Nicoline van der Sijs (Heerlen, 1955) is hoogleraar Historische taalkunde van het Nederlands in de digitale wereld aan de Radboud Universiteit Nijmegen. Ze studeerde Slavisch aan de Universiteit Utrecht, waar ze van 1975 tot 1990 als onderzoeker werkte. In 2001 promoveerde ze in Leiden. Tussen 1990 en 2010 publiceerde ze als freelance-onderzoeker ruim 25 boeken. In 2001 ontving ze de ANV-Visser-Neerlandiaprijs, in 2006 de Prins Bernhard Cultuurfonds Prijs. Sinds 2010 werkt ze bij het Meertens Instituut, o.a. als projectleider van het NWO-project Nederlab.

DE VOORTZETTING VAN DE HISTORISCHE TAALKUNDE MET ANDERE MIDDELEN



## **De voortzetting van de historische taalkunde met andere middelen**

*Rede uitgesproken bij de aanvaarding van het ambt van hoogleraar Historische taalkunde van het Nederlands in de digitale wereld aan de Faculteit der Letteren van de Radboud Universiteit Nijmegen op donderdag 27 maart 2014*

**door prof. dr. Nicoline van der Sijs**

Vormgeving en opmaak: gloed*communicatie*, Nijmegen  
Fotografie omslag: Bert Beelen  
Drukwerk: Van Eck & Oosterink  
Figuren in de tekst: Rob Zeeman

ISBN 978-90-9028087-5

© Prof. dr. N. van der Sijs, Nijmegen, 2014

Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar worden gemaakt middels druk, fotokopie, microfilm, geluidsband of op welke andere wijze dan ook, zonder voorafgaande schriftelijke toestemming van de copyrighthouder.

*Mijnheer de rector magnificus, geachte aanwezigen,*<sup>1</sup>

Een eeuw geleden, in 1915, schreef de Nijmeegse hoogleraar Jacques van Ginneken: '[we] zouden [...] met betrekkelijk weinig moeite: de wetenschap onzer Nederlandsche taal zóó hoog op kunnen stooten tusschen al hare concurrenten, dat het waarlijk een lust zou zijn het te zien.' Daarvoor was het wel nodig methodisch nieuwe wegen in te slaan, namelijk 'de nauwkeurigheid der exacte wetenschappen na te volgen: in de precisie van het getal.'<sup>2</sup> Statistiek heeft de toekomst, aldus Van Ginneken. Tél taalverschijnselen, adviseerde hij: 'En niet in het woordenboek natuurlijk, – daar staan immers allerlei termen die zelden of nooit gebruikt worden, met woordjes die minstens tienmaal in elk gesprek voorkomen, gelijkwaardig naast elkander; en de buigings- en vervoegingsvormen staan er meestal zelfs heelemaal niet in! – maar in een doorlopenden tekst, of liever nog uit den mond van een spreker, en het allerliefst uit de levende gesprekswisseling van het dagelijksch verkeer.'

Met deze aansporing tot corpusonderzoek was Van Ginneken zijn tijd vooruit. De uitwerking die hij in zijn artikel levert, moeten we – zoals bij veel van zijn onderzoek – met een stevig korreltje zout nemen: zo beschouwt hij de gemiddelde zinslengte en de 'gemiddelde variatie' in het aantal lettergrepen per zin als 'thermometer' voor 'het zieleleven en temperament' van 'onze grootere schrijvers'.<sup>3</sup> Met deze berekeningswijze liep hij vooruit op de leesbaarheidsformule die de Amerikaan Rudolf Flesch in 1948 ontwikkelde en die nog steeds is ingebouwd in het tekstverwerkingsprogramma Microsoft Word. Ook die formule berust op de zinslengte en het aantal lettergrepen per woord, maar berekent daarmee het leesgemak van een tekst. Eigenlijk rekende Van Ginneken dus niet het temperament van onze grote schrijvers uit, maar de leesbaarheid van hun teksten.

Sinds 1915 heeft de statistische of kwantitatieve methode binnen de taalwetenschap duidelijk terrein gewonnen, zij het aanzienlijk minder snel dan wetenschappers in de loop der jaren telkens verwachtten. Zo meende de lexicograaf Félicien de Tollenaere in 1963 dat men binnen maximaal vijf jaar met machines 'in staat zal zijn zeer snel een gedrukte tekst foto-elektrisch te lezen zoals hij is' – al voegde hij er voorzichtigheids-halve aan toe: 'Dit is echter vermoedelijk rijkelijk optimistisch.'<sup>4</sup> Dat was bepaald niet overdreven: pas sinds een jaar of vijftien jaar is de techniek zo ver ontwikkeld dat bibliotheken het tekstuele erfgoed op grote schaal kunnen digitaliseren, zoals we dat tegenwoordig noemen.

Ik citeer De Tollenaere niet om zijn voorspelling te ridiculiseren, maar om te laten zien hoe langzaam het integreren van de computer in de geesteswetenschappen gaat.

#### E-HUMANITIES

Ondanks de inspanningen van Hugo Brandt Corstius hield zich in de vorige eeuw slechts een select gezelschap onderzoekers bezig met de zogenoemde *computertaalkunde*, of met *alfa-informatica*.<sup>5</sup> Pas in het begin van de eenentwintigste eeuw gingen steeds



meer geesteswetenschappers de technische mogelijkheden enthousiast omarmen. Dat ging gepaard met een naamswijziging: sinds 2004 zoemt de term *digitale geesteswetenschappen* of *e-humanities* rond in de academische wereld.<sup>6</sup> De keuze voor een nieuwe naam was zeker ook een imagokwestie. Het doel was te benadrukken dat het geesteswetenschappelijk onderzoek centraal staat en computertechnologie er slechts de dienstmaagd van is.<sup>7</sup> Op die manier hoopte men dat meer geesteswetenschappers zich tot de nieuwe leer zouden bekeren. Want niet iedereen stond of staat te juichen bij het feit dat technologie in de geesteswetenschappen een steeds grotere rol krijgt.

In dat opzicht is er de laatste vijftig jaar minder veranderd dan je misschien zou verwachten. Want ook in 1963 achtte De Tollenaere het nodig geruststellende woorden te wijden aan het inzetten voor taalkundig onderzoek van wat hij noemde 'de elektronische rekenmachine, dikwijls ook, met een Engelse naam, *computer* genoemd.' 'Tegenover de automatisering [...] zijn drie houdingen denkbaar', aldus De Tollenaere: ten eerste een afwerende houding, waarbij men er 'a priori van uit[gaat], dat automatisering zou neerkomen op deshumanisering van de humanistische vakken. De techniek gevoelt men als een bedreiging.' [...] 'Een tweede geesteshouding wortelt in een soort van blind, mysticistisch geloof in de almacht der machine. Men gaat zich verbeelden dat een druk op de knop ons de schitterendste resultaten in de schoot werpt.' Als een ware dialecticus zag De Tollenaere nog een derde mogelijke houding die deze tegengestelde reacties in zich verenigt: 'Het is die waarbij men zich nuchter en zakelijk de vraag stelt: wat heeft de automatisering [...] te bieden? Wie deze laatste geesteshouding tot de zijne maakt, is niet bang voor de machine, doch evenmin blind voor haar beperkingen.'<sup>8</sup>

Dit is vijftig jaar geleden geschreven, toen een computer nog een met ponskaarten gevoed monster was dat een half huizenblok besloeg. Veel veranderd is er echter niet: de drie door De Tollenaere genoemde geesteshoudingen zijn anno 2014 nog alom te vinden. Zo manen NIOD-onderzoekers Hinke Piersma en Kees Ribbens in het laatste nummer van de *Bijdragen en Mededelingen betreffende de Geschiedenis der Nederlanden* tot kritische reflectie op digitaal onderzoek.<sup>9</sup> Heilig geloof in de almacht van de computer blijkt daarentegen uit de telkens herhaalde mantra dat de digitale geesteswetenschappen 'the next big thing' zijn.<sup>10</sup> Eenzelfde hooggestemd vertrouwen toont het rapport dat begin dit jaar verscheen naar aanleiding van het voornemen van de KNAW, de Amsterdamse universiteiten en IBM om een Center for Humanities and Technology op te richten, met een acroniem CHAT genoemd.<sup>11</sup> De samenstellers stellen dat de huidige automatische tekstanalyse al in staat is tot het herkennen van patronen en thema's, het extraheren van informatie en het analyseren van associaties. Volgens hen zijn computers inmiddels volwaardige cognitieve systemen die binnen afzienbare tijd in natuurlijke taal met mensen zullen communiceren en samenwerken.

Het rapport leest als een sciencefictionroman. Hier worden enkele stappen overgeslagen, althans in de ogen van een historisch taalkundige.<sup>12</sup> Want laten we wel wezen: met moderne teksten kan de computer momenteel aardig uit de voeten, maar bij

historische teksten vormt alleen al een simpel gegeven als de spellingvariatie een groot struikelblok.<sup>13</sup> Het einde van de Geesteswetenschappen 1.0, dat mijn Amsterdamse collega Rens Bod in 2012 inluidde, is voor historische teksten nog lang niet bereikt.<sup>14</sup>

De derde houding die De Tollenaere beschrijft, die van het nuchter de mogelijkheden maar óók de beperkingen onder ogen zien, spreekt mij dan ook het meest aan. Van beide – de mogelijkheden én de beperkingen – zal ik voorbeelden geven.

#### EEN DRAMA IN VIJF BEDRIJVEN

Geesteswetenschappers die zich bezighouden met het verleden – historisch taalkundigen, historisch letterkundigen en (kunst)historici – halen per definitie hun gegevens uit oude teksten, die immers de neerslag van de Nederlandse taal, literatuur, geschiedenis en cultuur vormen. Vroeger vergaarden de onderzoekers informatie door tijdrovende *close reading*. Tegenwoordig kunnen ze dankzij de computer enorme hoeveelheden digitale teksten snel doorzoeken. Dit *distant reading*<sup>15</sup> gaat echter bepaald niet probleemloos, want de digitalisering staat pas in de kinderschoenen en de huidige situatie werd al een ‘digitaal drama’ genoemd.<sup>16</sup>

Dat drama voltrekt zich als een klassiek treurspel in vijf bedrijven. In het eerste bedrijf is er een grote afwezigheid: een blauwdruk voor het digitaliseren van teksten. Instellingen stemmen nauwelijks met elkaar af welke teksten ze digitaliseren, met als gevolg overlappingsen én lacunes in het gedigitaliseerde materiaal. De digitale teksten worden bovendien op verschillende plaatsen en in verschillende bestandsformaten aangeboden, waardoor onderzoekers ze niet tegelijkertijd kunnen doorzoeken en analyseren. Stel je wilt reconstrueren hoe de discussie over *vrouwenkiesrecht* is verlopen tussen 1848, toen mannen voor het eerst rechtstreeks leden van de Tweede Kamer mochten kiezen, en 1919, toen ook vrouwen dit felbegeerde recht kregen. Je wilt weten of het debat door mannen of door vrouwen werd gestuurd, en welke argumenten pro en contra in de verschillende media en in de Tweede Kamer werden gebruikt. Bij de huidige stand van zaken is dergelijk onderzoek aan de hand van digitaal toegankelijke bronnen niet uitvoerbaar – ook niet via Delpher, het toegangsportaal van de wetenschappelijke bibliotheken.

In het tweede en derde bedrijf speelt de gebrekkige kwaliteit van de metadata en de data de hoofdrol. De metadata, zoals bibliografische gegevens, zijn niet of nauwelijks geüniformeerd. Daardoor laten de vindbaarheid en de toegankelijkheid van teksten nog veel te wensen over. De meeste data, de teksten, zijn op dit ogenblik door massadigitalisering ontsloten. Helaas maakt de optische tekenherkenning waarmee de oude teksten automatisch worden gelezen, nog veel fouten, waardoor de zoekresultaten onbetrouwbaar zijn. Als je Delpher mag geloven, dateert *televisie* van 1853, *fietsen* van 1676, en de eerste *tweet* van 1645.

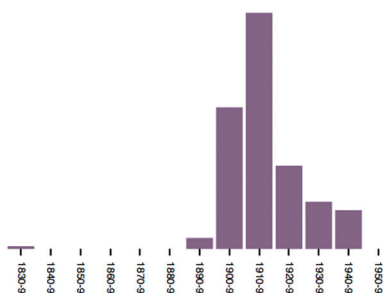
In het vierde bedrijf treden de tools op, de computerprogramma’s waarmee men de teksten analyseert. Die tools werken meestal slechts goed bij specifieke teksten, bijvoorbeeld alleen bij die uit de moderne tijd en in de huidige spelling. Tools worden

lang niet altijd in nauwe samenwerking met onderzoekers gebouwd, zodat het uiteindelijke resultaat regelmatig niet voldoet aan de behoefte van de onderzoekers. Vaak ontbreekt ook een userinterface, een gebruikersschil die de tool bruikbaar maakt voor niet-technici. Het gevolg van dit alles is dat slechts een klein deel van de ontwikkelde tools daadwerkelijk wordt gebruikt.<sup>17</sup>

#### NEDERLAB

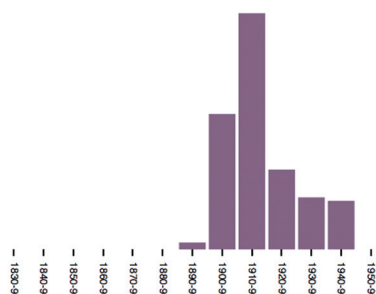
Een klassieke tragedie eindigt droevig, maar gelukkig is het digitale drama, om het op z'n zeventiende-eeuws te zeggen, een 'blijvend treurspel'. De positieve wending is dat onderzoekers en technici in 2013 de handen ineengeslagen hebben om samen de problemen te lijf te gaan. Afgelopen jaar zijn we begonnen met de constructie van een speciaal onderzoeksinstrumentarium voor de historische geesteswetenschappen, namelijk Nederlab. Nederlab zal een laboratorium, een werkplaats, gaan bieden voor onderzoek naar veranderingspatronen in de Nederlandse taal en cultuur. Het project wordt gedragen door alle belangrijke geesteswetenschappelijke onderzoeksinstellingen en faculteiten in Nederland, en wordt gefinancierd door NWO en KNAW.<sup>18</sup>

Via Nederlab kunnen onderzoekers op termijn alle gedigitaliseerde Nederlandstalige teksten van circa 800 tot nu op een centraal platform doorzoeken. Daarvoor worden de metadata van de verschillende dataleveranciers semi-automatisch aan elkaar gekoppeld. Op dit ogenblik staat er van Nederlab nog niet meer dan een voor buitenstaanders afgeschermd bètaversie. Maar daarin kan al het een en ander. Bij wijze van voorbeeld blijven we nog even bij het onderwerp *vrouwenkiesrecht*. Op figuur 1 ziet u de verdeling van de publicaties waarin hierover tussen 1840 en 1950 is gesproken, afgezet op een tijdsas. Het blijkt dat het onderwerp pas na 1890 echt gaat leven en dat er een piek is in de periode 1910-1919, natuurlijk niet toevalligerwijs rond het jaartal waarin het vrouwenkiesrecht werd ingevoerd.

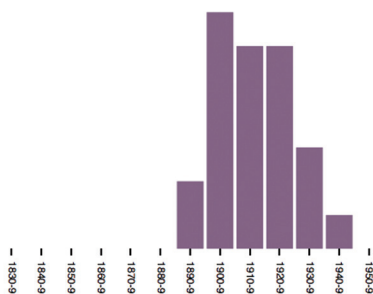


Figuur 1. Publicaties over vrouwenkiesrecht tussen 1840 en 1950

Op de twee volgende figuren worden de publicaties uitgesplitst naar het geslacht van de auteur.



Figuur 2. Mannen over vrouwenkiesrecht.

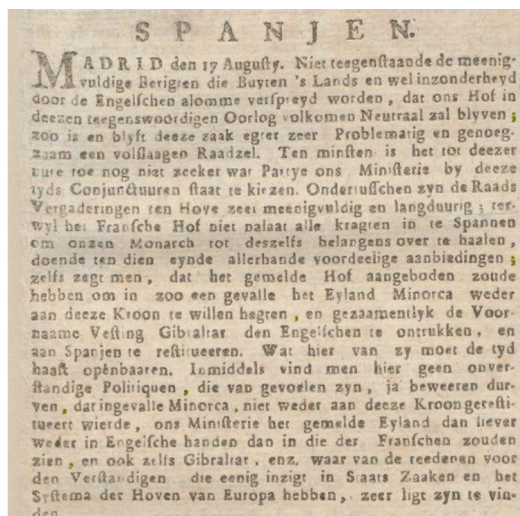


Figuur 3. Vrouwen over vrouwenkiesrecht.

Hieruit blijkt dat de bijdragen van mannen hetzelfde verloop kennen als dat van alle schrijvers gezamenlijk, met een piek tussen 1910 en 1919, terwijl vrouwen een decennium eerder dan de mannen piekten, en daarmee wellicht de discussie aanwakkerden die de mannen in de daaropvolgende periode voerden. Bovendien schreven vrouwen ook na de toekenning, in de jaren twintig, nog steeds veelvuldig over het onderwerp. De mannen hadden het er toen wel mee gehad.

We kunnen uit deze figuren niet aflezen hoe de bijdrage van de vrouwen aan de discussie zich verhoudt tot die van de mannen. Ook dat kunnen we binnen Nederlab inzichtelijk krijgen. Het blijkt dat slechts een schamele twaalf procent van de teksten waarin *vrouwenkiesrecht* voorkomt, van de hand van een vrouw is. Dat is nauwelijks meer dan het percentage publicaties dat in die periode *überhaupt* door vrouwen werd geschreven: dat bedraagt volgens de steekproef die Nederlab nu toestaat namelijk elf procent. Het onderwerp *vrouwenkiesrecht* deed vrouwen begin vorige eeuw dus niet massaal naar de pen grijpen.

Terwijl via Nederlab teksten makkelijker vindbaar zijn en beter te analyseren dankzij fijnmazige metadata, heeft ook Nederlab niet één ultieme oplossing voor het kwaliteitsgebrek van de data als gevolg van slechte optische tekenherkenning. Maar we doen er van alles aan. Leesfouten corrigeren we zoveel mogelijk met behulp van het door Martin Reynaert ontwikkelde programma TICCL.<sup>19</sup> Dit verbetert bijvoorbeeld automatisch woorden als *fchaap* en *fchip* in *schaap* en *schip*: de lange *s* uit het oude drukwerk is in deze woorden ten onrechte herkend als *f*. Bij teksten van voor 1900 lopen we daarbij echter wel tegen de technische grenzen aan, zoals blijkt uit het willekeurige krantenbericht uit 1756 in Figuur 4. Van de 243 woorden die het krantenbericht telt, zijn er na correctie nog steeds 62 onjuist gelezen, ofwel ruim één op de vier. Hier ligt een handschoen die de technici van CHAT hopelijk zullen oppakken.



Resultaat van de optische tekenherkenning:

MADRID den 17 Augusty. Niet tegenftaaode de meenigvuldige Bciigten die Buyten “\* Lands en wel inzonderheyd door de Engelfchen alomme verftpt.yd worden, dat ons Hof in deesen tšegenswoordigen Oorlog volkomen Neutraal zal blyven; 300 is en blyft deeze” zaak eg'er zeer Problematic en genoega,iam een volflaagen Raadael. Ten minden is het tot deezer tue to. nog niet zeker war Pattye ons Min'fterie by deeze tyds Conjunöuurn ftaat te kiezen. Ondenuffchen zyn dcšaads deringen ten Hov« zeci meeuigvuldig en langduung 5 rertiet Frarfch. Hof niet nalaat alle kragren in te Spannen cm onzen Monarch tot deszelfs belargens over te haaien, doende ten dien eynde allerhande voordeeüge aanbiedingen; zelfs zegt men, dat het gemelde Hof aangeboden zoude hebben om in zoo «en gevalle het Eyland Minorca weder aan daeze Kroon te willen hegren, en gezaamentlyk de Vooraaanie Vesting Gib>altat den Engelfchen te ontrukken, en aan Sparjen te redt uecren. Wat hier van zy moet de tyd haaft openbaaren. Icmidd.l» vind men hier geen onverftandige Folitiquen, die van gevoelen zyn, ja beweeren durven, dat ingevalle Minerea, niet weder aan deeze Kroon gereftitueerc wierde, ons Min'fterie het gemelde Eyland dan liever Weder in Engelfche handen dan in die der Fianfchen zouden zien, en ook ztlls Gibraltar. enz. waar van de reedenen voor den Verftatdigen die eenig inzicht in S aat» Zaaken en het Srftema der Hoven van Europa hebben,- zeer ligt zyn te vin\* den.

Na automatische correctie:

M A D R I D den 17 August. Niet teegenftaaode de menigvuldige Bciigten die Buyten ‘\* Lands en wel inzonderheid door de Engelfchen lamme verfpt.yd worden, dat ons Hof in Deense tRegenswoórdigen Oorlog volkomen Neutraal zal bóven ; 300 is en blafft Dieze zaak V6’er zeer Problematiek en genoegzaam een volslaagen Raadael. Ten minden is het tot dezer tue to. nog niet zeker war Patty ons ministerie by Dieze tyds Conjunöuurn aftast te kiezen. Ondenuffchen zyn dcßaads dringen ten Hov« zeci menigvuldig en langdurig 5 termiet fransch Hof niet nalaat alle kragen in te Spannen cm onzen Monarch tot deszelfs belagers over te haaien, doende ten dien Reyne allerhande voordelige aanbiedingen ; zelfs zegt men, dat het gemelde Hof aangeboden zoude hebben om in zoo «en gevalle het Wyland Minorca weder aan dazere Kroon te willen hegen , en gezaamentlyk de Voorgaande vesting Gibraltar den Engelfchen te ontrukken, en aan Spanje te redt euren. Wat hier van zy moet de tyd Haaften openbaarden. lcmidd.l» vind men hier geen onverstandige Folitiquen , die van gevoelen zyn, ja beweren durven , dat ingevallen mieren , niet weder aan Dieze Kroon gerestituteerd wierde, ons ministerie het gemelde Wyland dan liever Weder in Engelsch handen dan in die der Franfche zouden zien , en ook zetels Gibraltar . enz. waar van de redenen voor den Verftat-digen die eenig rinzig in S aat» Zaken en het Srftema der Hoven van Europa hebben,- zeer ligt zyn te vin\* den.

Figuur 4. Een nieuwsbericht uit de Opregte Groninger Courant van 10-9-1756, met het resultaat van de optische tekenherkenning en daarna het resultaat van automatische correctie: vet de gewijzigde woorden, onderstreept woorden die ook na correctie fout zijn gelezen.<sup>20</sup>

Daarnaast biedt de Nederlab-interface allerlei mogelijkheden om intelligent te zoeken, bijvoorbeeld met wildcards, want hoe slimmer onderzoekers kunnen zoeken, hoe beter ze de klippen van de slechte data kunnen omzeilen.<sup>21</sup> Dankzij het computationeel-historisch lexicon dat op het Instituut voor Nederlandse Lexicologie (INL) is ontwikkeld, hoeft men bij het zoeken ook niet meer alle historische woordvarianten te kennen en te definiëren: de computer vindt automatisch de 14 spellingvarianten van *regering* (Figuur 5) en de 26 vervoegde vormen van *lopen* (Figuur 6).

De teksten worden in Nederlab bovendien voorberekt, zodat onderzoekers veel gericht naar resultaten kunnen zoeken. Aan iedere woordvorm wordt door het programma Frog, dat door onder anderen Antal van den Bosch werd ontwikkeld, informatie toegevoegd over woordenboekvorm, woordsoort, uitgang en morfologische geleding; in technische termen: de teksten worden gelemmatiseerd, getagd en geparsed. Daarbij maakt de computer ook een verschil tussen het telwoord *acht* en het homonieme zelfstandige naamwoord *acht* ‘aandacht’.



Alle woorden: "regering" (woordvar)

**Opslaan**

---

**Verfijn zoekresultaat**

Toon resultaten, gevonden:

☒ In tekst

☐ In metadata

regeering: 48848

regeeringe: 819

regeeringen: 4813

regeeringh: 49

regeeringhe: 61

regering: 11829

regeringe: 547

regeringen: 2271

regeringhe: 98

regheringhe: 14

regieringe: 214

regierungen: 27

regieringh: 43

regieringhe: 107

sorteren op: Datum ▼ stijgend ▼

---

**Omzien in verwondering** Collectie: Digitale Bibliotheek voor de Nederlandse Letteren

---

**Jehan Froissart's Cronyke van Vlaenderen. Deel 1**  
Collectie: Digitale Bibliotheek voor de Nederlandse Letteren

---

**M/F** Collectie: Digitale Bibliotheek voor de Nederlandse Letteren

---

**Jehan Froissart's Cronyke van Vlaenderen. Deel 4**  
Collectie: Digitale Bibliotheek voor de Nederlandse Letteren

---

**De historie van Partinoples, grave van Bleys**  
proza (ridderroman)  
Publicatie: ca. 1250  
Auteur: anoniem Historie van Partinoples, grave van Bleys, De (-)  
Collectie: Digitale Bibliotheek voor de Nederlandse Letteren

Figuur 5. Spellingvarianten van regering.

Alle woorden: "lopen" (woordvaria)

**Opslaan**

---

**Verfijn zoekresultaat**

Toon resultaten, gevonden:

☒ In tekst

☐ In metadata

eloopen: 69

gelopen: 5575

gelopen: 3486

ghelopen: 247

ghelopen: 2

laupe: 10

liep: 22661

liepen: 11248

liept: 241

loepen: 554

loop: 34702

loope: 641

loopen: 21527

lopend: 2353

lopende: 6114

lopenden: 575

loopen: 156

loope: 16

loope: 4

loopen: 20

loopen: 1

loops: 2402

loopt: 29093

lope: 837

lopen: 13472

lopende: 3890

sorteren op: Datum ▼ stijgend ▼

---

**Getijdenboek**  
proza (gebed(en), non-fictie/theologie, vertaling)  
Publicatie: 14de eeuw  
Auteur: Geert Grote (Deventer, 1340-Deventer, 1384) Collectie: Digitale Bibliotheek voor de Nederlandse Letteren

---

**Alles loopt mij tegen**  
poëzie (liedblad / straatlied)  
Publicatie: 19de eeuw  
Auteur: anoniem Alles loopt mij tegen (-)

... Alles **loopt** mij tegen. Wijze: Ik zal mij wreken, valsche meid. Mijn vrienden en vriendinnen al, 't ...

Collectie: Digitale Bibliotheek voor de Nederlandse Letteren

---

**Omzien in verwondering**

... magnetische macht over mij gehad. Toen ik nauwelijks kon **lopen**, **liep** ik, als men mij een ogenblik aan het ...

... vadet officier werd, en ik weet nog hoe glorieus wij kinderen met hem over straat **liepen** in zijn nieuwe ...

... praten, met mannen die in overhemdsmouwen en met vrouwen die overdag met krulspelden in het haar **liepen** ...

... ons zonder fantasie, zoals het hoort, en Ti en ik **liepen** tot zeven, acht jaar toe met kort geknipte ...

... stap vast **liep** in de ijzere greep van de paranoia. Wie nu uit dit alles zou afleiden, dat ik een ...

Collectie: Digitale Bibliotheek voor de Nederlandse Letteren

Figuur 6. Vervoege vormen van lopen.

Nooit eerder zijn in de Lage Landen op zó grote schaal zóveel verschillende tekstbestanden taalkundig verrijkt. Bij deze voorbewerkingen geldt echter een belangrijk voorbehoud: automatische verrijking werkt vooralsnog heel matig op historische teksten.<sup>22</sup> Het programma Frog is wel zelflerend en kan dus voor oudere teksten worden getraind. Dat willen we in de toekomst binnen Nederlab gaan doen.

#### EEN HISTORISCHE THESARUS

Al deze voorbewerkingen hebben als doel vat te krijgen op de enorme vormvariatie van het Nederlands. Voor onomasiologische variatie, dus het bestaan van uiteenlopende benamingen voor één en hetzelfde begrip, hebben we nog geen hulpmiddel. En dat, terwijl de meeste begrippen in de loop van de tijd verschillende meer of minder synonieme benamingen hebben gehad. Stel dat je de geschiedenis van de zuigelingenverzorging in de Lage Landen wil schrijven. Dan moet je alle termen kennen waarmee men in het verleden baby's aanduidde.

babbaard, bakerkind, bébé, bloed, boontje, doezel, dreumel, dreumes, dreutel, hokkebrok, jongere, jongkind, keutel, kindje, kleintje, kleuter, koter, kriel, kroost, kruimel, mormeldier, pappaard, peukel, peuter, peuzel, pieter, poeter, pop, pruts, puk, schreeuwer, schreeuwlelijk, tot, uk, wezel, wicht, worm, zuigeling

Figuur 7. Een aantal synoniemen voor 'baby' door de eeuwen heen (in gemoderniseerde spelling).

Voor textmining van historische teksten is een digitale historische thesaurus een conditio sine qua non. Op een enkel deelterrein wordt hieraan al gewerkt. Zo wordt in het NWO-project Time Capsule van Toine Pieters een historische thesaurus van medicinale planten gerealiseerd. Momenteel buig ik me samen met het INL over de vraag hoe we een historische thesaurus semi-automatisch uit de bestaande historische woordenboeken van het Nederlands kunnen destilleren.

Na dit globale overzicht van wat Nederlab op termijn te bieden heeft, zal ik nu enkele concrete voorbeelden geven van onderzoek dat voor het eerst mogelijk wordt dankzij de nieuwe digitale onderzoeksmethoden. Daarbij beperk ik mij tot casussen op het gebied van mijn leeropdracht: de historische taalkunde.

#### DE STATENVERTALING: ARCHAÏSCH OF NIET?

In mijn boek *Taal als mensenwerk. Het ontstaan van het ABN* uit 2004 heb ik onder andere de stelling naar voren gebracht dat de grote invloed van de Statenvertaling op de vorming van de standaardtaal een mythe is. De Statenvertaling leverde weliswaar, zo betoogde ik, diverse nieuwe woorden en uitdrukkingen op, maar op de spelling en op de meeste



grammaticale kwesties had ze geen blijvende invloed, omdat ze bij verschijnen in 1637 al archaïsch was en afweek van wat toen gebruikelijk was. Zo gebruikten de Statenvertalers een dubbele ontkenning (*en ...niet*), en drukten ze het voegwoord 'en' uit met *ende*: *ende ghy en sult niet liegen*. Ze volgden hiermee het middeleeuwse systeem, terwijl in het dagelijkse taalgebruik van begin zeventiende eeuw al werd gezegd *en ghy sult niet liegen*. Schrijvers als Hooft en Cats waren al eerder overgestapt van een tweeledige naar een eenledige ontkenning.<sup>23</sup>

In weerwil van de mijns inziens ijzersterke argumenten bleek niet iedereen het met mij eens. Toen ik dat boek schreef, had ik nog niet de beschikking over een corpus en tools. Nu wel, en nu is de vraag: hoe toets je of de Statenvertaling bij verschijnen archaïsch was of niet? Een eenvoudige manier is om te kijken naar het percentage verbogen naamvallen. In het Middelnederlands behoorden de naamvallen tot het levende taalgebruik, maar in de vroegmoderne tijd werden ze in de Hollandse spreektaal steeds meer weggelaten. Het verschijnsel heet deflexie en er is veel over geschreven, recent nog door Joop van der Horst.<sup>24</sup> In de schrijftaal bleef men daarentegen veel waarde hechten aan de naamvallen: alle grammatica's uit die tijd wijdden er lange passages aan, omdat men meende dat in een volmaakte en rijke taal verschillende functies of betekenissen moesten worden weergegeven door verschillende vormen. Jammer genoeg voldeed het Nederlands daar niet aan. Daarom hielpen de spraakkunstenaars het een handje door kunstmatig nieuwe vormen te scheppen. Aan die drang tot verbetering danken we tot op heden onder andere het verschil tussen *hen* en *hun*.

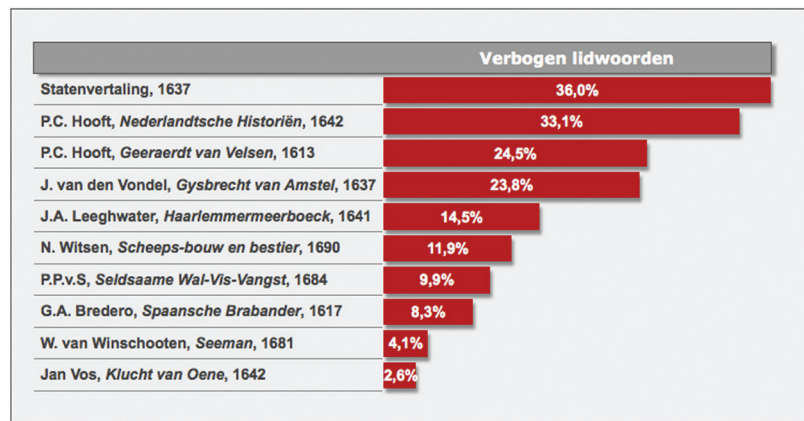
Ook de Statenvertalers zaten met de deflexie in de maag: voordat ze met de vertaling begonnen, hadden ze regels opgesteld juist rond het gebruik van de naamvallen, en hun werk werd nog eens grondig door revisoren nagekeken voordat het naar de zetter ging. Mijn hypothese is dat de Statenvertaling meer verbogen naamvallen bevat dan andere tekstsoorten uit de zeventiende-eeuwse Republiek, doordat de vertalers de voorschriften van de spraakkunstenaars streng navolgden en een verheven toon nastreefden. Andere teksten, met name die van laagopgeleiden, zullen dichter liggen bij de dagelijkse spreektaal en dus minder verbogen vormen bevatten.

Inmiddels kan deze hypothese worden getoetst, want vrijwilligers hebben de complete tekst van de Statenvertaling gedigitaliseerd.<sup>25</sup> Ook andere teksten uit diezelfde tijd zijn inmiddels digitaal beschikbaar, zodat we voor het eerst het aandeel verbogen vormen in de Statenvertaling kunnen vergelijken met dat in andere teksten uit dezelfde periode. Bij de telling heb ik mij beperkt tot de lidwoorden: de onverbogen vormen *de*, *het*, *een* versus de verbogen vormen *des*, *der*, *den* en *eenes*, *eener*, *eenen*, *eene*. Naamvallen werden natuurlijk ook uitgedrukt in bijvoeglijke naamwoorden en bezittelijke en aanwijzende voornaamwoorden, maar die kunnen we alleen tellen als de teksten zijn gelemmatiseerd en getagd, wat met zeventiende-eeuwse teksten nog niet goed gaat. Er staan echter zo veel lidwoorden in de teksten – gemiddeld is negen procent van alle woorden in de tekst een lidwoord – dat deze een betrouwbare indicator vormen voor de berekening van

het percentage verbogen vormen.<sup>26</sup>

De deflexie had in de aanloop tot de zeventiende eeuw al fors toegeslagen. In veertiende-eeuwse oorkonden, zo berekende Piet van Reenen, is maar liefst 63 procent van de lidwoorden verbogen. Anderhalve eeuw later, in 1556, bevat het *Nieuwe Testament* van Jan Utenhove nog 57 procent verbogen lidwoorden. In de Statenvertaling uit 1637 blijkt nog maar 36 procent van de lidwoorden verbogen te zijn. Zo gezien heeft de deflexie dus ook in de Statenvertaling doorgewerkt.

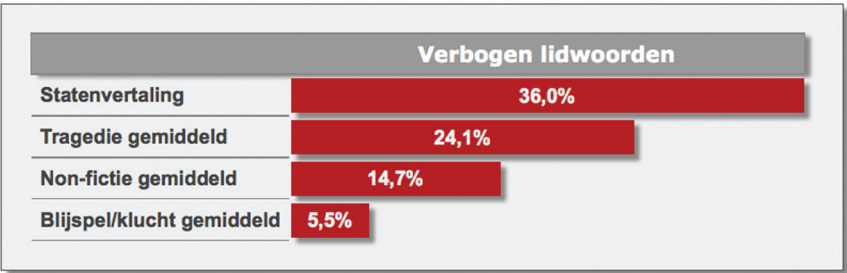
Maar als we de Statenvertaling vergelijken met andere teksten uit de zeventiende eeuw, komt het behoudende karakter ervan duidelijk naar voren. Figuur 8 laat in aflopende volgorde het percentage verbogen lidwoorden zien in literaire, zakelijke en drama-tische teksten.



Figuur 8. Het percentage verbogen lidwoorden in een aantal literaire, zakelijke en dramatische teksten uit de zeventiende-eeuwse Republiek.

Binnen al deze genres heeft de Statenvertaling het hoogste percentage verbogen lidwoorden. Waar in de Statenvertaling nog meer dan één op de drie lidwoorden verbogen wordt, is dat in de klucht van Jan Vos nauwelijks meer dan één op de veertig!<sup>27</sup>

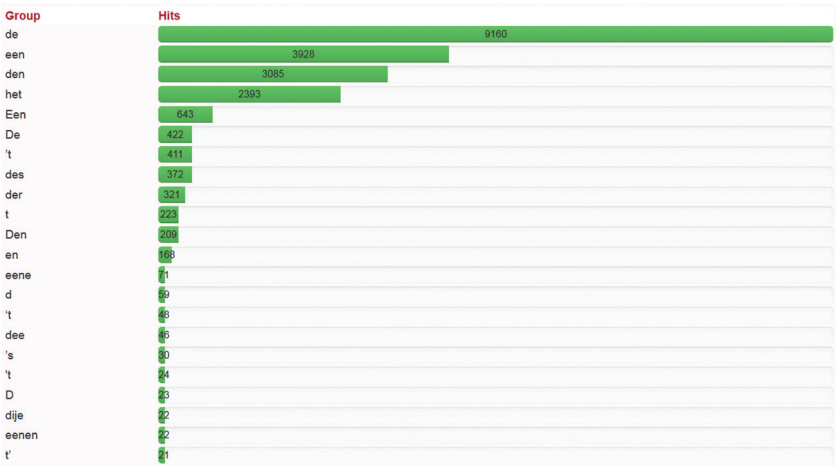
Uit Figuur 9 blijkt dat het percentage verbogen lidwoorden samenhangt met het tekstgenre. Na de Statenvertaling bevatten de tragedies de meeste verbogen lidwoorden: ruim 24 procent. Zakelijke teksten hebben er gemiddeld bijna 15 procent, en blijspelen of kluchten zelfs maar 5,5 procent.<sup>25</sup> Blijspelen lagen vermoedelijk het dichtst tegen de alledaagse spreektaal; dat blijkt ook uit het feit dat het totale percentage lidwoorden in deze teksten laag ligt, rond de 7 procent.



Figuur 9. Het percentage verbogen lidwoorden per tekstsoort.

Tot slot is het interessant te zien hoeveel verbogen vormen er voorkomen in brieven. We hebben uiteraard geen geluidsopnamen uit de zeventiende eeuw, maar we nemen vooralsnog aan dat brieven van laaggeletterden het beste beeld geven van de toenmalige spreektaal – een nog beter beeld dan blijspelen, die immers door de handen van zettters zijn gegaan. Sinds kort beschikken we over een groot aantal transcripties van zeventiende-eeuwse brieven afkomstig van door de Engelsen gekaapte schepen. Die transcripties zijn vervaardigd door vrijwilligers bij het Meertens Instituut en binnen het door Marijke van der Wal geleide project Brieven als buit.

Kenmerkend voor de spreektaalige blijspelen bleek dat het totale percentage lidwoorden laag lag (7 procent), en dat het percentage verbogen vormen nóg lager lag (5,5 procent). Als de hypothese dat brieven het dichtst bij de spreektaal liggen klopt, verwachten we dat de cijfers daarin nog onder die van de blijspelen blijven. Voor wat



Figuur 10. Lidwoorden in gekaapte brieven tussen 1600 en 1700.

betreft het totale percentage lidwoorden komt die verwachting uit: de brieven bestaan gemiddeld voor slechts 5,5 procent uit lidwoorden. De tweede verwachting, dat het percentage verbogen lidwoorden eveneens laag ligt, komt verrassend genoeg niet uit: in de door het Meertens Instituut gedigitaliseerde brieven tel ik 17,6 procent verbogen vormen. In de collectie Brieven als buit, die door het INL is gelemmatiseerd en getagd, ligt het aantal verbogen lidwoorden nog hoger, op 21 procent. Hoe kunnen we dat hoge aantal verklaren?

Als we de gegevens nauwkeuriger bekijken, dan blijkt dat veel van die verbogen lidwoorden voorkomen in een tijdsbepaling (*den 1 februari, den tijd, des avonts*), in aanspreekvormen als *den heer*, en in andere vaste formules. Dergelijke vaste verbindingen zeggen niets over iemands gebruik van naamvallen, tenminste ik veronderstel dat niemand in deze zaal wil beweren dat we in het Nederlands nog naamvallen gebruiken omdat we bijvoorbeeld *ten enen male of de heer des huizes* schrijven. Het aantal verbogen lidwoorden in de gekaapte brieven weerspiegelt dus maar ten dele de spreektaal. Judith Nobels, die eind vorig jaar op de brieven promoveerde, kwam tot een vergelijkbare conclusie. Zij vond dat in de zeventiende eeuw de naamvals -s opvallend vaak voorkwam in gevallen waarin je eerder een omschrijving met *van* zou verwachten. Die gevallen bleken echter grotendeels vaste literaire en religieuze formules te zijn, zoals *Gods genade, de hant des heeren of brenger deeses*.<sup>29</sup> En die religieuze formules, u raadt het al, stamden voor een groot deel uit de Statenvertaling, waarvan de archaïsche grammaticale vormen in de zeventiende en achttiende eeuw kennelijk een sacrale beleveniswaarde behielden.

Uit de zojuist besproken gegevens blijkt in elk geval zonneklaar dat de Statenvertaling meer verbogen lidwoorden bevat dan de andere onderzochte tekstsoorten uit dezelfde periode. Daarmee wordt het archaïsche karakter van de Statenvertaling althans in dit opzicht bevestigd. Maar er blijven nog genoeg vragen over. Bijvoorbeeld of er regionale verschillen bestaan in de teruggang van verbogen naamvallen. Algemeen wordt ervan uitgegaan dat ze in de Zuidelijke Nederlanden makkelijker voortleefden dan in het noorden. Het feit dat het *Schilderboek* van de in het West-Vlaamse Meulebeke geboren Karel van Mander uit 1604 ruim 40 procent verbogen lidwoorden bevat, meer dus dan de Statenvertaling, lijkt daarop te wijzen. Intrigerend is ook dat in de loop van de achttiende eeuw in de Republiek het aantal verbogen vormen weer enigszins toeneemt: dat bleek zowel in een zakelijke tekst als in de gekaapte brieven. Betekent dit dat de naamvalsregels in navolging van grammaticavoorschriften in de loop van de tijd weer strikter zijn toegepast? Het lijkt erop dat taalveranderingen, die zich volgens de klassieke hypothese van Anthony Kroch in een vast tempo verbreiden, in geschreven taal kunnen worden vertraagd – zij het slechts tijdelijk.<sup>30</sup>

De uiteindelijke fundamentele vraag die we willen beantwoorden is wat de oorzaak of de motor van deflexie is. Joop van der Horst legt in zijn recente studie *Taal op drift* een algemeen causaal verband met maatschappelijke veranderingen, maar noemt geen spe-

cifieke oorzaak. De meeste onderzoekers zoeken die in taalcontact<sup>31</sup> en de invloed van tweedetaalverwervers.<sup>32</sup> Voor tweedetaalverwervers zijn uitzonderingen in de grammatica lastig en daarom systematiseren ze de grammaticale regels: ze maken de grammatica regelmatiger. Dat is precies wat er bij deflexie gebeurt: de moeilijk leerbare naamvals-vormen verdwijnen. Als deflexie inderdaad wordt veroorzaakt door taalcontact, zou er een correlatie moeten bestaan tussen het aantal taalveranderingen in een bepaalde periode en de migratie van dialectsprekers of de immigratie van buitenlanders naar de Lage Landen. De verwachting is: hoe meer verplaatsingen, hoe meer taalveranderingen. Er bestaan momenteel nog geen databases met gegevens over enerzijds verplaatsingen en anderzijds taalveranderingen, maar zodra die er zijn, moeten we ze aan elkaar kunnen linken: als de AIVD het kan, kunnen wij het ook.

#### NEDERLANDSE CONTACTVARIËTEITEN

Mijn tweede voorbeeld van onderzoek met behulp van digitale technologie behelst het Nederlands van andere continenten: Suriname, Indonesië, de Nederlandse Antillen, Zuid-Afrika. Van dergelijk buitengaats Nederlands weten we zeker dat er sprake is van veel verplaatsingen, veel taalcontact en ook veel taalverandering. In de verschillende gebieden zijn Nederlandse contactvariëteiten of etnolecten ontstaan: Surinaams-Nederlands, Antilliaans-Nederlands, Indisch-Nederlands en Petjoh, een mengtaal van Maleis en Nederlands. Die contactvariëteiten onderscheiden zich in uitspraak, woordvorming, woordenschat en grammatica van het Europees-Nederlands. Van die varianten bestaan globale beschrijvingen, maar systematisch onderzoek naar de onderscheidende elementen ontbreekt.

Zouden we daar slimme technieken voor in kunnen zetten? Kan de computer bijvoorbeeld neologismen opsporen? Als test hebben technici van het Meertens Instituut op mijn verzoek met behulp van software een vergelijking gemaakt tussen een Nederlandstalig krantencorpus uit Nederland en een uit het toenmalige Nederlands-Indië, beide uit het jaar 1900.<sup>33</sup> Dit leverde een lijst van woorden op waarvan de frequentie in de twee bestanden statistisch significant verschilde. De woorden die alleen of vooral in Indische kranten voorkwamen maar nauwelijks in het moederland, zijn potentiële nieuwvormingen in het Indisch-Nederlands. De lijst bevat leenwoorden als *pasar* en *totok*, waarvan al bekend is dat ze uit Indië stammen. Daarnaast bevat de lijst ook heel gewone Nederlandse samenstellingen waarachter men nooit een Indische connectie heeft vermoed. Van die woorden heb ik in grotere tekstbestanden de oudste voorkomens bekeken. Daarbij bleek bij verschillende woorden dat ze daadwerkelijk in Indië zijn ontstaan, bijvoorbeeld *ontvang-avond* en *spreekdag*, die oorspronkelijk gebruikt werden voor de momenten dat een bestuurder in Indië bezoek ontving. Ook woorden als *pandgoederen*, *postsluiting*, *sluitdag*, *venduloaal* en *vervaldag* hebben een Indische herkomst.<sup>34</sup>



Figuur 11. Het voorkomen van postsluitingen en sluitdagen in kranten uit 1900.

Het automatisch vergelijken van twee bestanden is dus een vruchtbare methode om neologismen op te sporen. In de toekomst wil ik deze toepassen op een veel omvangrijker tekstbestand. Daarbij wil ik kijken of de methode nader kan worden uitgewerkt door voort te bouwen op de *Stable Lexical Marker Analysis*, een algoritme dat de Leuvense onderzoekers Dirk Spielman en Dirk De Hertog specifiek hebben ontwikkeld om verschillen tussen het Nederlands-Nederlands en het Belgisch-Nederlands op te sporen.<sup>35</sup>

Taalcontact leidt niet alleen tot nieuwvormingen, maar ook tot veranderingen in grammaticale regels – zoals ik al opmerkte naar aanleiding van het verdwijnen van verbogen naamvallen. Uit onderzoek naar Engelse contactvariëteiten weten we dat die verandering vaak zit in de distributie of de frequentie van een grammaticale constructie.<sup>36</sup> Zo geeft het Surinaams-Nederlands de toekomstige tijd bij voorkeur weer met het werkwoord *gaan*: *u gaat het niet vinden* of zelfs *u gaat het niet gaan vinden*. Ook het Standaardnederlands en Nederlandse dialecten, met name het Vlaams, gebruiken *gaan* wel voor de toekomstige tijd, maar minder vaak en in andere contexten dan het Surinaams-Nederlands.

Welke syntactische bijzonderheden kent het Surinaams-Nederlands nog meer? Die vraag hebben we met behulp van de computer proberen te beantwoorden. Daarvoor hebben Katrien Depuydt en Jesse de Does van het INL een modern krantencorpus uit Suriname en uit Nederland syntactisch laten ontleden door het programma Alpino van Gertjan van Noord. Daarna hebben ze gekeken welke syntactische verschijnselen in het ene corpus statistisch significant vaker voorkwamen dan in het andere.

Het belangrijkste verschil dat naar voren kwam, was dat in de kranten uit Nederland vaker twee hoofdzinnen zonder markeringswoord op elkaar volgen, waar het Surinaams-Nederlands kiest voor een hoofdzin gevolgd door een bijzin. Zo lees je in een Nederlandse krant: *Rotterdam is voor haar van groot belang: het is de eerste stad waar alleen nog met de chipkaart kan worden gereisd*, terwijl een Surinaamse krant een voegwoord als *omdat* invoegt: **omdat** *het de eerste stad is waar alleen nog met de chipkaart kan worden gereisd*. Op dat moedertaalsprekers van het Nederlands minder behoefte hebben aan voegwoor-

den om verbanden tussen zinnen te leggen, maar het verschil is niet wereldschokkend. Waarschijnlijk zijn de krantencorpora te veel volgens specifieke normen geredigeerd, waardoor ze onvoldoende het dagelijkse taalgebruik reflecteren.

Erik Tjong Kim Sang van het Meertens Instituut heeft Alpino vervolgens enkele Surinaams-Nederlandse en Nederlandse romans automatisch laten ontleden. Hieruit kwamen enkele bijzonderheden van het Surinaams-Nederlands naar voren. Ten eerste het gebruik van een nabepaling die begint met *zo*, bijvoorbeeld: *hoe je 'em daar zag, zó snel*, of *jullie vertellen direct wat jullie van iemand denken, zó vrijpostig*, waar wij zouden zeggen: *hij ging zó snel* en *jullie zijn zó vrijpostig*. Ten tweede het gebruik van *komen* als hulpwerkwoord, wat Surinaams-Nederlandse zinnen oplevert als: *Gusta kwam niet zichzelf vertonen! of Wanneer regen zich liet komen vallen*.<sup>37</sup>

Tijdens deze tests bleek dat Alpino moeite heeft met het ontleden van niet-standaardtalige zinnen en die zinnen dan probeert te transformeren in standaardtaal. Hierdoor strijkt het programma bijzonderheden in de oorspronkelijke tekst glad in plaats van ze als bijzonderheden te markeren. Dit probleem moeten we nog oplossen.

Hoewel dit nog maar proefmetingen zijn, laten ze toch al zien dat corpusgebaseerd onderzoek naar contactvariëteiten veelbelovende vooruitzichten biedt.<sup>38</sup> Een vergelijking van zoveel mogelijk contactvariëteiten – oude en jonge – zal aan het licht brengen welke taalkundige kenmerken de variëteiten gemeen hebben. Dat zijn waarschijnlijk grammaticale vereenvoudigingen door tweedetaalverwerving. Van sommige van die kenmerken veronderstellen onderzoekers dat ze in de toekomst onderdeel gaan uitmaken van het Standaardnederlands. Zo wordt vaak beweerd dat het Nederlands onder invloed van allochtonen gaat naar één woordgeslacht en één lidwoord, en constructies als *een mooie boek* en *de meisje die* dus de toekomst hebben.<sup>39</sup> Omdat deze veranderingen binnen de deflexie-beweging vallen – het Engels en Afrikaans gingen ons immers voor met het enkele lidwoord *the* en *die* – zal deze voorspelling ooit wel uitkomen.

Van andere taalveranderingen die in contactvariëteiten plaatsvinden, is het moeilijker te voorspellen of ze beklijven of niet. Daarvoor hebben we inzicht nodig in de voorwaarden waaronder taalcontact leidt tot *blijvende* taalveranderingen. We weten bijvoorbeeld dat in alle contactvariëteiten de neiging bestaat het voornaamwoord *het* weg te laten (*trek je niet aan*), *er* te vervangen door *daar* (*ik heb zin daarin*) of weg te laten (*wat wil je aan doen, ik heb twee*), en in hoofdzinnen, bijzinnen en vraagzinnen een vaste woordvolgorde te gebruiken (*hij zegt dat je moet dit niet geloven*). Toch heb ik nog nooit een serieuze taalkundige horen voorspellen dat *er* het loodje zal leggen, of dat in de toekomst de woordvolgorde in Nederlandse hoofd- en bijzinnen identiek zal zijn. Terwijl dit goed denkbaar is, want beide verschijnselen komen voor in Nederlandse dialecten.<sup>40</sup>

Veel – misschien wel de meeste – taalveranderingen in contactvariëteiten zijn niet doorgedrongen tot het Standaardnederlands. Waarom niet? Is dat een kwestie van tijd, of zijn er structurele factoren die sommige taalveranderingen tegenhouden? Digitaal onderzoek naar Nederlandse contactvariëteiten kan hier nader licht op werpen. Dat



onderzoek wordt nu nog bemoeilijkt doordat er maar weinig corpora van contactvariëteiten van het Nederlands beschikbaar zijn. Daar wil ik in de toekomst samen met collega's van Nederlandse en Belgische universiteiten verandering in brengen.

#### VERANDERINGEN IN WOORDVORMINGSPATRONEN

Het laatste voorbeeld van digitaal onderzoek dat ik u wil voorleggen, draait om veranderingen in woordvormingspatronen. Ik ben al jaren gefascineerd door de vraag welke veranderingen er hebben plaatsgevonden in de manier waarop Nederlandse afleidingen zijn gevormd: woorden met een voorvoegsel als *be-* in *bedenken* of met een achtervoegsel als *-ig* in *groenig*.<sup>41</sup> Het aantal woorden dat met een bepaald voor- of achtervoegsel kan worden afgeleid, is in principe onbeperkt (hoewel aan restricties gebonden), maar de voorraad voor- en achtervoegsels zelf is eindig en kan voor de verschillende periodes van het Nederlands worden vastgesteld.

Tien jaar geleden liet ik in mijn proefschrift zien dat er in de zestiende en zeventiende eeuw structurele veranderingen optraden in de manier waarop afleidingen werden gevormd: achtervoegsels zoals *-nis* in *duisternis* en *-st* in *kunst* verloren hun productiviteit en werden vervangen door andere, nieuwe achtervoegsels.<sup>42</sup> Het lijkt geen toeval dat juist in die periode de standaardtaal vorm kreeg, maar we weten eigenlijk heel weinig over het hoe en waarom van deze veranderingen. Behalve Matthias Hüning en Marijke Mooijaart hebben tot dusver weinig onderzoekers aandacht besteed aan de geschiedenis van Nederlandse afleidingen. Dat heeft niets te maken met gebrek aan wetenschappelijke belangstelling, maar er stonden praktische bezwaren in de weg: er bestaat geen geschikt onderzoekscorpus, en de programma's die woorden ontleden, zoals het al eerder genoemde Frog, werken alleen op modern Nederlands.

Wel beschikken we over moderne gegevens en gegevens uit veertiende-eeuwse oorkonden, die taalkundig zijn verrijkt. Dankzij die verrijking kunnen we bekijken welke achtervoegsels in de veertiende eeuw voorkwamen en met welke basiswoorden ze werden gecombineerd. Als voorbeeld neem ik de achtervoegsels waarmee een zelfstandig naamwoord van een bijvoeglijk naamwoord werd afgeleid. Daarvoor gebruikte men maar liefst zeven achtervoegsels: *-e*, *-de* of *-te*, *-heid*, *-dom*, *-nis*, *-schap* en *-ing*. Dikwijls werd een en hetzelfde basiswoord gecombineerd met verschillende achtervoegsels: van *beter* werden *beterheid*, *beterschap*, *beternis* en *betering* afgeleid, van *lang* maakte men *lange* en *lengte*, en van *vrij* leidde men *vrije*, *vrijheid* en *vrijdom* af.

Na de middeleeuwen werden nieuwe afleidingen vrijwel alleen nog gevormd met het achtervoegsel *-heid*; uit de twintigste eeuw stammen bijvoorbeeld *abstractheid* en *fitheid*. De variatie in achtervoegsels die de basiswoorden in de veertiende eeuw kenden, is in de loop van de tijd fors ingeperkt. De meeste basiswoorden hebben één of hoogstens twee achtervoegsels behouden, althans in de standaardtaal: we kennen wel *zekerheid*, maar niet *zekernis*. Als er twee achtervoegsels bewaard zijn gebleven, hebben die vaak verschillende betekenissen gekregen, vergelijk *liefde* en *liefheid*.



Dit alles maakt nieuwsgierig: hoe verliepen de veranderingen in de distributie van de achtervoegsels, en waardoor kunnen ze verklaard worden? Om een begin van een antwoord te vinden, moeten we in ieder geval meer weten over de frequentie en regionale verspreiding van de achtervoegsels in de veertiende eeuw. Het ligt immers voor de hand dat achtervoegsels die altijd maar weinig voorkwamen of in een klein gebied werden gebruikt, eerder zullen verdwijnen dan achtervoegsels die over het hele Nederlandse taalgebied veel voorkwamen. Figuur 12 toont de frequentie van de achtervoegsels in veertiende-eeuwse oorkonden.

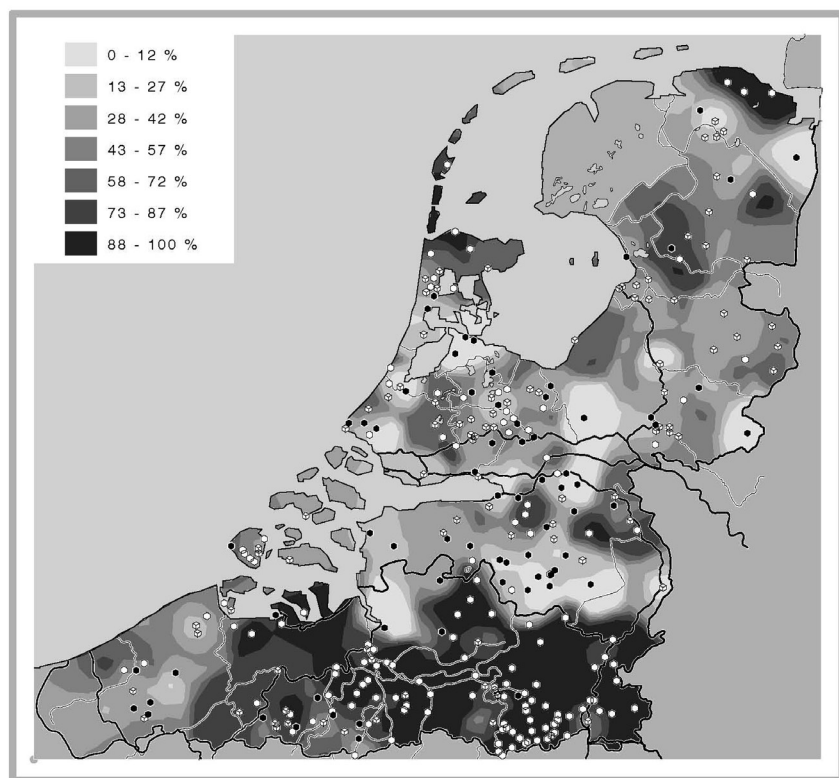
-heid	1875 keer in 56 basiswoorden	<i>o.a. achterstalligheid, afstelligheid, bedachtigheid, behendigheid, beterheid, edelheid, eenheid</i>
-e	496 keer in 16 basiswoorden	<i>o.a. brede, erfelijke, heilige, zieke</i>
-de/-te	400 keer in 11 basiswoorden	<i>liefde, dikte, gemeente, gewoonte, grofte, grootte, hoogte, lengte, meente, volte, ziekte</i>
-dom	524 keer in 4 basiswoorden	<i>eigendom, vrijdom, vroeddom, wijsdom</i>
-nis	157 keer in 7 basiswoorden	<i>beternis, klaarnis, mindernis, vastenis, waarnis, zekernis, zwaarnis</i>
-schap	94 keer in 3 basiswoorden	<i>beterschap, vroedschap, waarschap</i>
-ing	64 keer in 3 basiswoorden	<i>goeding, lozing, wijzing</i>

Figuur 12. De frequentie van de achtervoegsels -heid, -e, -de/-te, -dom, -nis, -schap en -ing in veertiende-eeuwse oorkonden.

Het achtervoegsel *-heid* komt verreweg het vaakst voor. Figuur 13 bevat een kaart die Piet van Reenen heeft getekend met de regionale verbreiding van het achtervoegsel *-heid* in de veertiende eeuw.<sup>43</sup> In Figuur 14 ziet u de verbreiding van de samengenomen achtervoegsels *-e/-de/-te*: in deze periode werd het achtervoegsel *-e* geleidelijk vervangen door *-de* of *-te*.

Uit de kaarten krijgen we een indruk van de regionale herkomst van de achtervoegsels: de afleiding *-heid* lijkt te stammen uit het Oost-Vlaams en zuidelijke Brabants en Limburgs, want daar vormt dit achtervoegsel het dichtste netwerk, terwijl *-e* vooral voorkomt in het oosten, met name de regio Gelderland en noordoostelijk Limburg. De verbreiding van de achtervoegsels is deels complementair: in Noord-Holland, Overijssel en vooral het zuiden heeft *-heid* de overhand, in het oosten *-e*. Maar de verbreiding van de

heid in oorkonden versus oorkonden zonder



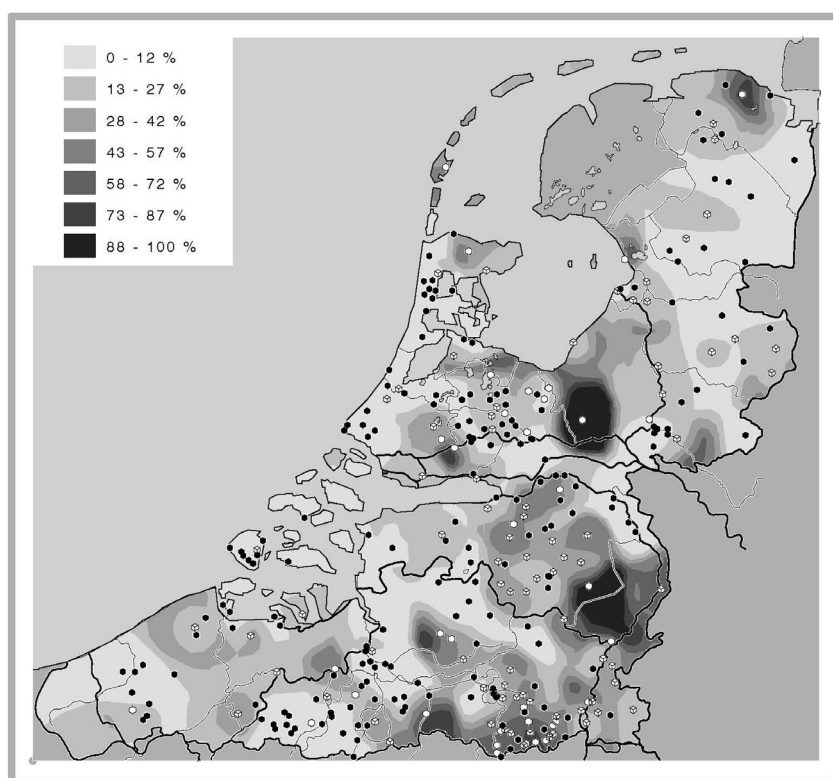
programma: e wattel vu amsterdam

31-12-13

18:37

Figuur 13. De regionale verdeling van het achtervoegsel -heid in de veertiende eeuw.

e (of de of te) s in oorkonden versus oorkonden zonder



Figuur 14. De regionale verbreiding van de achtervoegsels -e/-de/-te in de veertiende eeuw.

achtervoegsels in de veertiende eeuw laat, in combinatie met de frequentie, eigenlijk vooral een vooraankondiging zien van het feit dat *-heid* nu de normale vorm is in de standaardtaal.

De gegevens uit de veertiende eeuw vormen slechts een momentopname en daarmee alleen kunnen we de veranderingen in de woordvorming uiteraard niet definitief verklaren. Naast frequentie en regionale variatie spelen factoren mee als de klank en de betekenis van de achtervoegsels: een achtervoegsel dat te weinig onderscheidend is, is minder levensvatbaar, en volgens veel onderzoekers verklaart dat waarom het achtervoegsel *-e* is vervangen door *-te*, zodat we een 'vast gebruik' niet meer *gewone* noemen maar *gewoonte*.<sup>44</sup> Daarnaast kunnen sociolinguïstische factoren misschien verklaren waarom het Nederlands zoveel geleende achtervoegsels kent, denk aan *-ier* in *scholier*, *-ij* in *razernij* of *-iteit* in *stommititeit*.<sup>45</sup>

Om verder onderzoek te doen naar de geschiedenis van Nederlandse afleidingen ga ik samen met de Belgische onderzoekers Marijke De Belder en Vincent Vandeghinste subsidie aanvragen bij NWO-FWO.

#### TOT SLOT

Aan de oproep die Van Ginneken honderd jaar geleden deed om taalverschijnselen in corpora te gaan tellen, wordt inmiddels ruimschoots gehoor gegeven. Ik heb u vandaag enkele mogelijkheden van dat soort onderzoek laten zien, maar nog lang niet alles: zo hebben we het nog helemaal niet gehad over speciale statistische software.<sup>46</sup> Een belangrijke toepassing waar ik nog even bij wil stilstaan, is het op zinvolle wijze aan elkaar linken van data uit verschillende bronnen. Dat is meer een kwestie van dataverrijking dan van onderzoek doen, maar met zo'n verrijking kunnen we wel nieuwe antwoorden vinden op oude vragen. Zo ben ik met streektaaldeskundigen in Nederland en Vlaanderen bezig een elektronische Woordenbank van de Nederlandse Dialecten op te zetten. Hierin worden zoveel mogelijk oude en jonge dialectwoordenboeken opgenomen, en de gegevens worden gekoppeld aan materiaal uit eerder gehouden dialectenquêtes, de Meertens Kaartenbank met ruim 30.000 taalkaarten en de Etymologiebank met 40 etymologische bronnen. Op basis van dit materiaal kunnen we bijvoorbeeld te weten komen hoe stabiel dialectgrenzen zijn, en welke invloed regionalisering en standaardisering op de dialecten hebben.<sup>47</sup>

Veel onderzoekers en studenten staan te trappelen om het digitale onderzoeksterrein te betreden, maar weten niet hoe ze hun onderzoeksvragen zo moeten stellen dat de computer bij de beantwoording kan helpen. Alle Nederlandse humaniorafaculteiten organiseren cursussen en workshops e-humanities. Het ontbreekt nog aan een toegankelijk handboek voor de digitale geesteswetenschappen, waarin geesteswetenschappelijke onderzoeksvragen centraal staan, en waarin wordt beschreven met welke data en technieken die vragen het best beantwoord kunnen worden, welke voorbewerkingen daarvoor noodzakelijk zijn, en voor welke valkuilen men moet waken. Een dergelijk handboek wil

ik in de loop van volgend jaar met collega's van deze en andere universiteiten publiceren.

Ik hoop dat de onderzoeksvoorbeelden u een indruk hebben gegeven van de mogelijkheden die de digitale wereld biedt aan de historische taalkunde. Het is mijn overtuiging dat de geesteswetenschappen dankzij technische innovaties aan het begin van een nieuw tijdperk staan. Daarom ben ik erg blij met mijn mooie leeropdracht, waarin ik de kans krijg historische taalkunde en digitale onderzoeksmethoden te combineren. In mijn ogen is de digitale historische taalkunde – met een variatie op Carl von Clausewitz – een voortzetting van het onderzoek met andere middelen: de fundamentele onderzoeksvragen blijven dezelfde, namelijk het achterhalen welke patronen en wetmatigheden ten grondslag liggen aan taalveranderingen. De middelen – de digitale onderzoeksmethoden – helpen bij de zoektocht naar antwoorden. Dat er in het digitale onderzoek nog veel problemen zijn te overwinnen, zoals ik óók heb laten zien, is alleen maar een aanmoediging om er des te meer in te investeren, niet alleen financiële middelen, maar ook inventiviteit.

#### DANKWOORD

Tot slot een woord van dank. Om te beginnen ben ik Antal van den Bosch, Paula Fikkert, Ans van Kemenade en Margit Rem bijzonder dankbaar voor het feit dat zij mij naar Nijmegen hebben uitgenodigd: vanaf de eerste werkdag voelde ik me hier helemaal thuis. Het is een voorrecht om een bijdrage te mogen leveren aan het goed doortimmerde programma van de opleiding Nederlandse Taal en Cultuur. Dat het programma zo hecht is, is vooral de verdienste van het dagelijks bestuur, bestaande uit Margit Rem en Wilbert Spooren. Het college van bestuur van de Radboud Universiteit Nijmegen en het bestuur van de Faculteit der Letteren, met name Paul Sars en Theo Engelen, dank ik voor het in mij gestelde vertrouwen.

Ik profiteer veel van de discussies met de collega's van het Centre for Language Studies, met studenten, en met de promovendi die ik begeleid: Hans Beelen, Jan Luitzen, Theo Muller, Anneke van der Putte en Ewoud Sanders. Heel stimulerend is de samenwerking met Radboud-onderzoekers van diverse pluimage, zoals Peter-Arno Coppen, Ad Foolen, Roeland van Hout, José Sanders, Pieter Muysken en de leden van zijn onderzoeksgroep. De taaltechnologen Antal van den Bosch, Iris Hendrickx, Henk van den Heuvel, Nelleke Oostdijk en Martin Reynaert ben ik dankbaar voor het geduld waarmee ze hun grote technische kennis met me delen.

Ook contacten buiten Nijmegen zijn voor mij heel waardevol: met de Meertens-collega's en in het bijzonder Hans Bennis, die mij ervan heeft overtuigd dat het Meertens Instituut inmiddels het tijdperk van *Het bureau* ver achter zich heeft gelaten, met de technici, redacteurs en onderzoekers van Nederlab, met collega's van Huygens ING, INL, Fryske Akademy, en met een groot aantal Vlaamse taalkundigen. Het allerbelangrijkste is natuurlijk de steun van vrienden als Hans Beelen, Jaap Engelman, Joep Kruijsen, Ewoud Sanders en René van Stipriaan. Aparte vermelding verdient mijn promotor Piet van Sterkenburg. Tot slot noem ik mijn moeder en andere familie-

leden, die met verbazing hebben gezien dat je met een bul in de slavistiek de kost kunt verdienen.

Mijn leeropdracht, waarin e-humanities centraal staan, is bij uitstek interdisciplinair. E-humanities doe je niet alleen: het vakgebied stimuleert tot samenwerking, tussen onderzoekers van verschillende geesteswetenschappelijke disciplines, en tussen onderzoekers en technici. Daarbij worden universitaire grenzen geslecht, en kennis en wetenschappelijke ideeën worden gedeeld. Dit alles wordt gesymboliseerd in mijn toga. Hij ziet eruit als een echte Radboud-toga, maar schijn bedriegt: het is een omgekatte vu-tabbaard, die bijna 40 jaar lang bij academische plechtigheden is gedragen door Elrud Ibsch. Dankzij Anja de Feijter en Andrea Kunne gaat de toga een nieuw leven tegemoet in een andere universiteit en een andere discipline. Als dat niet het slechten van grenzen is!<sup>48</sup>

Na een lang verblijf in de echte wereld ben ik sinds kort weer teruggekeerd in de academische moederschoot – niet alleen hier in Nijmegen maar ook in Amsterdam. Het is een bijzondere en ook leerzame ervaring. They do things differently there. De bureaucratische overlegstructuur is wennen, maar ik geniet van de inhoudelijke discussies, en ik hoop me daarin nog lang te kunnen mengen.

*Ik heb gezegd.*

## NOTEN

- 1 Voor hun nuttige opmerkingen dank ik Sjeff Barbiers, Hans Beelen, Jaap Engelsman, Joep Kruijsen, Felix van de Laar, Piet van Reenen, Margit Rem, Ewoud Sanders, René van Stipriaan en Ton van der Wouden.
- 2 J. van Ginneken (1915), 'De statistiek en de taalwetenschap', in: *De Nieuwe Taalgids* 9, 65-95, citaten op p. 65, 68.
- 3 Idem p. 84.
- 4 F. de Tollenaere (1963), *Nieuwe wegen in de lexicologie*, Amsterdam, p. 101.
- 5 Zie voor de geschiedschrijving van het vak Leonoor van der Beek (2011), *Van rekenmachine tot taalautomaat*, Groningen.
- 6 In dat jaar verscheen bij Blackwell het overzichtswerk *A Companion to Digital Humanities*. De benaming *digitale geesteswetenschappen* is opmerkelijk: er bestaat immers niet zoiets als *digitale natuurwetenschappen* of *digitale sociologie*.
- 7 Nicoline van der Sijs (2013), 'Het nut van computerlezen', in: *NRC Handelsblad*, Wetenschapsbijlage 6 april, p. 2.
- 8 De Tollenaere 1963: 123-124.
- 9 Hinke Piersma & Kees Ribbens (2013), 'Digital Historical Research. Context, Concepts and the Need for Reflection', in: *BMGN – Low Countries Historical Review* 128, 4, 78-102.
- 10 Deze karakterisering van de digital humanities stamt uit 2009 en is afkomstig van de Amerikaanse hoogleraar William Pannapacker, zie: <http://bit.ly/1oFF5hW>.
- 11 *Meaning and Perspectives in the Digital Humanities. A White Paper for the Establishment of a Center for Humanities and Technology (CHAT)*, eds David Millen & Sally Wyatt.
- 12 In de vs woedde rond de jaarwisseling een vergelijkbare discussie: John Markhoff 'Brainlike Computers, Learning From Experience' in: *The New York Times* 29-12-2013 ([nyti.ms/1gsaTSo](http://nyti.ms/1gsaTSo)) versus Gary Marcus 'Hyping Artificial Intelligence, Yet Again' in: *The New Yorker* 1-1-2014 ([nry.kr/1ewXs1z](http://nry.kr/1ewXs1z)).
- 13 Zie bijvoorbeeld Hans van Halteren & Margit Rem (2013), 'Dealing with Orthographic Variation in a Tagger-Lemmatizer for Fourteenth-Century Dutch Charters', in: *Language Resources and Evaluation* 47, 4, 1233-1259.
- 14 Rens Bod (2012), *Het einde van de Geesteswetenschappen* 1.0, Amsterdam.
- 15 De term is gemunt door literatuuronderzoeker Franco Moretti van The Stanford Literary Lab.
- 16 Zie René van Stipriaan (2010), 'De nieuwe bibliotheek', in: *De Gids* 173, 326-334; 'Het digitale drama' in *NRC Handelsblad*, Wetenschap, 10 en 11 sept. 2011, met reacties op: <http://www.den.nl/nieuws/bericht/3229>.
- 17 Dat is vaak ook een geldkwestie. Voor de ontsluiting van data en de aanpassing van tools van het Nederlandstalige Referentiecorpus SoNaR heeft CLARIN onlangs geld uitgetrokken; het is te hopen dat hetzelfde gebeurt voor vergelijkbare applicaties, denk aan Adelheid, Cornetto, DutchSemCor, Alfalab, etc.
- 18 Zie <http://www.nederlab.nl/>.
- 19 Martin Reynaert (2010), 'Character Confusion Versus Focus Word-Based Correction of Spelling and OCR Variants in Corpora', in: *International Journal on Document Analysis and Recognition* 14, 173-187.
- 20 De krant is te vinden op [bit.ly/1fVXRgl](http://bit.ly/1fVXRgl). Soms heeft de automatische correctie een Verschlimmbesserung uitgevoerd (*alomme* > *lamme*). Eigenlijk is ook een verandering van *meenigvuldige* in *menigvuldige* onjuist, want het doel van automatische correctie is te komen tot een exacte weergave van de oorspronkelijke bron, en niet een omspelling naar de moderne spelling. Dergelijke veranderingen heb ik niet meegerekend; als je dat wel doet, kom je ná correctie op een foutenpercentage van 32,5%.

- 21 Zie *Eerste hulp bij e-Onderzoek* van Ewoud Sanders: <http://www.ewoudsanders.nl/zoeken.html>.
- 22 Dat probleem kennen overigens andere talen ook: zo schreef de Amerikaanse computerlinguïst Mark Liberman onlangs dat zelfs voor het lemmatiseren van modern Engels geen enkel werkelijk betrouwbaar programma bestaat (zie <http://bit.ly/1mnFyXz>).
- 23 Nicoline van der Sijs (2004), *Taal als mensenwerk: het ontstaan van het ABN*, Den Haag, pp. 534-537.
- 24 Joop van der Horst (2013), *Taal op drift. Lange-termijnontwikkelingen in taal en samenleving*, Amsterdam.
- 25 Zowel de eerste druk uit 1637 als de herziene uitgave van 1657 – die wat betreft het naamvallengebruik identiek zijn. De teksten staan op [www.dbnl.org](http://www.dbnl.org) en [www.bijbelsdigitaal.nl/](http://www.bijbelsdigitaal.nl/).
- 26 Ter vergelijking: in het hedendaagse *Corpus Gesproken Nederlands* is het aantal lidwoorden in voorgelezen geschreven teksten 9% en in vrije spraak 6,2%. Dat er in de spreektaal minder lidwoorden worden gebruikt, is te verklaren – zo bleek onder andere uit het proefschrift van G.J. Bouma (2008), *Starting a Sentence in Dutch. A Corpus Study of Subject- and Object-Fronting*, Groningen – doordat daarin vaker een voornaamwoord wordt verkozen boven een volle zelfstandignaamwoordgroep: *ik zag de man* wordt in de spreektaal meestal: *ik zag hem*.
- 27 De *Nederlandse Historiën* van Hooft vallen op vanwege het hoge aantal verbogen vormen. In deze geschiedschrijving volgde Hooft bij wijze van stijl oefening de Romeinse schrijver Tacitus na, en dat is te merken aan het gebruik van verbogen vormen. Daarmee deed hij naar het schijnt zijn eigen stijl geweld aan, want in het treurspel *Geeraert van Velsen* van dezelfde auteur ligt het percentage verbogen lidwoorden maar liefst een kwart lager. Aan de andere kant van het spectrum valt het maritieme woordenboek van Wigardus van Winschooten uit 1681 op: hierin is slechts 4,1 procent van de lidwoorden verbogen. Dat is opmerkelijk, want de geleerde Winschooten was leraar Latijn en hij keek dus niet op een naamval meer of minder. Maar Winschooten was een verre voorloper van Multatuli, want elders pleitte hij er met zoveel woorden voor om de Nederlandse taal te schrijven zoals men hem spreekt. En hij bracht die opvatting blijkens het zeer geringe aantal verbogen vormen moeiteloos zelf in praktijk. Zie verder: Hans Beelen, Ingrid Biesheuvel & Nicoline van der Sijs (2011), *Seeman. Maritiem woordenboek van Wigardus à Winschooten*, Zutphen, pp. 18-19.
- 28 Ook in de toneelstukken *Geeraert van Velsen* en *De Spaansche Brabander* is het totale percentage lidwoorden laag: ruim 6%. Dat komt overeen met het percentage in het moderne *Corpus Gesproken Nederlands*. De frequentie van lidwoorden is dus in de laatste drie eeuwen in het Nederlands niet veel veranderd.
- 29 Judith Nobels (2013), *(Extra) Ordinary Letters: A View from Below on Seventeenth-Century Dutch*, Utrecht, hfd 9. Zie voor 18de-eeuwse brieven: Tanja Simons (2013), *Ongekend 18e-eeuws Nederlands: Taalvariatie in persoonlijke brieven*, Utrecht.
- 30 Anthony S. Kroch (1989), 'Reflexes of Grammar in Patterns of Language Change', in: *Language Variation and Change* 1, 199-244. Zie ook S.A. Tagliamonte (2012), *Variationist Sociolinguistics: Change, Observation, Interpretation*, Oxford. Een belangrijke toevoeging aan Krochs model in: G.J. Postma (2010), 'The impact of failed changes', in: C. Lucas, S. Watts, A. Breitbarth & D. Willis (eds) (2010), *Continuity and Change in Grammar*, Amsterdam, pp. 269-302.
- 31 De Engelse sociolinguïst Peter Trudgill heeft in *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*, Oxford, 2011 laten zien dat kleine, geïsoleerde talen grammaticaal complexer zijn dan talen waarvan de sprekers meer contact hebben met andere talen.
- 32 Fred Weerman (1993), 'The Diachronic Consequences of First and Second Language Acquisition: the Change from OV to VO', in: *Linguistics* 31, 903-931; Wouter Kusters (2003), *Linguistic Complexity. The Influence of*



- Social Change on Verbal Inflection*, Utrecht; H.J. Bennis (2010), 'Darwin en taal; de stem van de evolutie', in: *vakTaal* 1/2, 19-21; Ad Foolen (2012), 'Zeker: voor een deel, terug naar Van Ginneken!', in: *Voortgang. Jaarboek voor de neerlandistiek* 30, 277-287.
- 33 Ik dank de technici Matthijs Brouwer, Hennie Brugman, Matthijs Droës, Marc Kemps-Snijders, Maarten Marx, Martin Reynaert, Erik Tjong Kim Sang, Rob Zeeman en Junte Zhang voor hun hulp.
- 34 Zie verder: Nicoline van der Sijs (2013), '*Sluitdagen en afpakschuur*: de Indisch-Nederlandse woordenschat', in: *De tienduizend dingen. Feestbundel voor Reinier Salverda*, Leeuwarden, pp. 389-402.
- 35 D. De Hertog, K. Heylen & D. Speelman (2012), 'Stable Lexical Marker Analysis: A Corpus-based Identification of Lexical Variation', in: Soares da Silva, A., A. Torres & M. Gonçalves (eds), *Pluricentric Languages: Linguistic Variation and Sociocognitive Dimensions*, Braga, pp. 61-74.
- 36 Zh. Bao (2010), 'Must in Singapore English', in: *Lingua* 120, 1727-1737. Bao toont aan dat in Singapore-Engels het modale werkwoord *must* vooral als deontisch werkwoord wordt gebruikt, en in Brits-Engels vooral epistemisch. Dat onderscheid in gebruik blijft zonder een gedigitaliseerd corpus onzichtbaar. Met dank aan Suzanne Aalberse.
- 37 C.M. de Kleine (1999) spreekt in haar dissertatie *A Morphosyntactic Analysis of Surinamese Dutch as Spoken by the Creole Population of Paramaribo*, Suriname (NY, pp. 156-157) van 'modaal' *komen*, en geeft twee anderssoortige voorbeelden daarvan in het Surinaams-Nederlands. Zie voor een beschrijving van het gebruik van *komen* ook: Hans Broekhuis & Norbert Corver (te verschijnen), *Syntax of Dutch: Verbs and verb phrases*, Amsterdam.
- 38 Zo is hieraan een aparte sectie gewijd in het standaardwerk *Language and Space: An International Handbook of Linguistic Variation* (2013), red. Frans Hinskens en Johan Taeldeman, Berlin.
- 39 O.a. in Jan Stroop (red.) (2003), *Waar gaat het Nederlands naartoe? Panorama van een taal*, Amsterdam; zie b.v. de bijdrage van Cornips en De Rooij.
- 40 Zie Sjef Barbiers, Hans Bennis e.a. (2005-2008), *Syntactische Atlas van de Nederlandse Dialecten*, 2 delen, Amsterdam, met voorbeelden in commentaardeel I, p. 25 (o.a. *Is nog soep?*) en commentaardeel II, p. 47, 50, 55 (o.a. *Jan vindt dat je moet dit niet geloven*). Evenmin wordt voorspeld dat het voornaamwoord *het* zal verdwijnen, terwijl daarvan in ieder geval in vaste verbindingen al aanzetten zijn te vinden: het Standaardnederlands heeft immers de constructie *fiíís goed* overgenomen uit het Surinaams-Nederlands en *klopt!* uit het Indisch-Nederlands (zie: J. van Donselaar (2003), 'fiíís goed', in: *Onze Taal* 72, 6, 148; Jan Stroop: <http://nederl.blogspot.nl/2013/05/klopt.html>). Uit corpusonderzoek blijkt dat de bewuste uitdrukkingen, met weglating van *het* of *dat*, het eerst en meest frequent in de contactvariëteiten voorkomen; ze zijn daarin waarschijnlijk ontstaan. Maar het weglaten van *dat* is ook in het Standaardnederlands productief in de zogenaamde topic-drop constructie (bijvoorbeeld *Heb ik al gezien* of *Kan.*), wat verklaart waarom de constructies gemakkelijk zijn overgenomen.
- 41 Zie Ariane van Santen & Nicoline van der Sijs (2006), 'Een historisch-morfologische database', in: *Taal en Tongval* themanummer 19, 153-168.
- 42 Nicoline van der Sijs (2001), *Chronologisch woordenboek van het Nederlands. De ouderdom en herkomst van onze woorden en betekenissen*, Amsterdam, hoofdstuk 3.1.
- 43 Kaarttechniek Evert Wattel, zie Evert Wattel & Pieter van Reenen (2010), 'Probabilistic maps', in: Alfred Lameli, Roland Kehrein, Stefan Rabanus (eds), *Language and Space, An International Handbook of Linguistic Variation*, Berlin, Volume 2: Language Mapping, Chapter 25, 495-505, 2501-2505.

- 44 A. van Loey (1970), *Schönfelds historische grammatica van het Nederlands*, Zutphen, pp. 234-235.
- 45 Ook het feit dat sommige achtervoegsels zijn voortgekomen uit zelfstandige woorden met een eigen betekenis, kan in hun verdere ontwikkeling een rol spelen; dat geldt bijvoorbeeld voor *-dom*, *-schap* en *-heid* (zie Nicoline van der Sijs (samensteller) (2010), *Etymologiebank*, op <http://etymologiebank.nl/>). Voor verschijnselen als grammaticalisatie en lexicalisatie zie Geert Booij (2002), *The Morphology of Dutch*, Oxford; Geert Booij & Ariane van Santen (1998), *De morfologie van het Nederlands*, Amsterdam; R.H. Baayen (1989), *A Corpus-Based Approach to Morphological Productivity. Statistical Analysis and Psycholinguistic Interpretation*, PhD thesis, VU Amsterdam.
- 46 Voor voorbeelden van toepassing in de literatuurwetenschap zie de oratie van Karina van Dalen-Oskam (2013), *De stijl van R*, Amsterdam.
- 47 Sceptici die menen dat de ontwikkelingen te snel gaan om vast te leggen in een handboek, zouden nog eens *Computer-taalkunde* van Hugo Brandt Corstius moeten herlezen: het boek dateert uit 1978 en wie het nu ter hand neemt, ziet dat er veel is veranderd maar ook dat veel problemen nog steeds niet zijn opgelost, zeker als het historische teksten betreft. Een ander inspirerend voorbeeld is *Eerste Hulp bij e-Onderzoek* van Ewoud Sanders, waarvan 125.000 exemplaren onder alfastudenten zijn verspreid.
- 48 Zie over de toga en andere rituelen: Matthijs Dicke, Paul van de Laar & Joop Visser (red.) (2013), *Ambitie en identiteit. Van Nederlandsche Handels-Hoogeschool tot Erasmus Universiteit Rotterdam, 1913-2013*, pp. 249-270; Willem Frijhoff 'Academische eer en wetenschappelijke reputatie, historisch beschouwd', in: W.Th.M. Frijhoff e.a. (1996), *Dies natalis Erasmus Universiteit Rotterdam, 8 November 1996*, Rotterdam, pp. 5-20.

